心理科学进展 2022, Vol. 30, No. 10, 2303–2320
Advances in Psychological Science

https://doi.org/10.3724/SP.J.1042.2022.02303

人工智能辅助的自闭症早期患者的筛查与诊断*

袁玉琢[1] 骆 方[2]

([1]中国基础教育质量监测协同创新中心; [2]北京师范大学心理学部, 北京 100875)

摘 要 自闭症谱系障碍(Autistic Spectrum Disorders, ASD)的症状早在婴幼儿期就会显现, 越早发现, 越早干预, 治疗效果越好。传统自闭症早期筛查与诊断在评估方法、流程上存在局限, 无法满足大规模筛查和诊断需求。随着人工智能技术的快速发展, 使用智能化方法进行自闭症早期大规模无感筛查与诊断逐渐成为可能。近 10 年间, 国内外对自闭症智能化识别方法的探索在经典任务行为、面部表情和情绪、眼动、脑影像、运动控制和运动模式、多模态 6 个领域积累了丰富的研究成果。未来研究应围绕构建国内自闭症早期智能医学筛查与诊断体系, 开发针对婴幼儿患者的筛查工具, 构建融合多模态数据的自闭症婴幼儿智能化识别模型, 建立结合脑影像技术的自闭症精细化诊断方法等方面来开展。

关键词 自闭症谱系障碍, 自闭症早期筛查与诊断, 自闭症智能化识别, 人工智能, 多模态数据
分类号 R395

1 引言

根据美国精神医学学会出版的第五版精神障碍诊断和统计手册(the Diagnostic and Statistical Manual of Mental Disorders-5th edition, DSM-5), 自闭症谱系障碍(Autism spectrum disorder, ASD)也被称为孤独症, 被定义为一种由神经系统发育失调导致的广泛发育障碍疾病(Hodges et al., 2020), 以社交(沟通)障碍和重复刻板的行为及兴趣为主要临床表现, 多见于儿童。

据世界卫生组织报道的 2019 年流行病学调查估计, 全世界每 160 名儿童就有 1 名罹患自闭症, 呈逐年上升的趋势。美国疾病控制与预防中心报告, 2016 年美国的自闭症发病率为 1/68, 至 2020 年美国的自闭症发病率已经达到 1/59 (Maenner et al., 2020)。在中国, 受限于起步晚和专业诊断人员的稀缺, 尚无权威的儿童自闭症流行病学数据。2014 年首次发布的《中国自闭症儿童发展状况报告》提到中国儿童自闭症患病率约为 1%, 2020 年对中国 6~12 岁儿童自闭症流行病学的首次全国性评估研究显示, 自闭症患病率为 0.70% (Zhou et al., 2020)。根据《中国孤独症(自闭症)教育康复行业发展状况报告Ⅲ》推算, 我国自闭症患者超过 1000 万, 其中 0~12 岁患者超过 200 万, 以每年近 20 万的速度增长。可见, 过去被视为罕见症的自闭症, 目前已经位居我国幼儿残疾发病率第二位, 仅次于智力障碍。

自闭症难以治愈、伴随终生, 不仅给患者个人造成痛苦, 给家庭和社会带来的负担也逐渐凸显。2015 年, 美国根据自闭症患病率发布了自闭症社会经济投入估算报告, 报告指出, 自闭症引发的医疗卫生成本、非医疗卫生成本和生产力合计全国 GDP 的 0.994%~2.009%, 预计到 2025 年, 将上升至 0.982%~3.6% (Liu et al., 2015)。我国有调查表明, 自闭症儿童的抚养成本(19582.4 元)明显高于智力障碍儿童(6391 元)与肢体残疾儿童(16410 元) (Dawson et al., 2018)。可见高发的自闭症对社会造成了极大的负担, 急需科学方法介入缓解其症状, 提高个人生活能力, 降低家庭和社会负担。

自闭症越早发现, 越早干预, 预后效果越好(Matson et al., 2008)。尤其婴幼儿的神经系统可塑性较高, 给予及时、适当的早期干预可以提高患者的适应能力和认知能力(徐云, 杨健, 2014)。但

收稿日期: 2021-07-26
* 国家自然科学基金联合基金项目(批准号: U1911201)。
通信作者: 骆方, E-mail: luof@bnu.edu.cn

是，父母往往在患儿 2~3 岁时才能发现明显的发育异常和行为表现，患儿从半岁到 2 岁间的轻微甚至是明显的异常往往因父母的经验不足而被忽略。此外，自闭症的诊断主要依据医生的经验，从父母发现儿童异常再到确诊又耗费很长的时间，缺乏便捷和客观有效的诊断手段。因此，开展针对婴幼儿的大规模自闭症早期无感筛查、发现风险案例后及早进入诊断流程是非常必要的。近 10 年，计算机视觉技术、语音技术、深度学习等人工智能和大数据挖掘技术已被有效应用于心理健康测评、自动化医疗诊断、疾病干预和康复领域，为婴幼儿自闭症的筛查和诊断带来了取得重大突破的可能性。将人工智能技术用于自闭症自动化、精细化筛查诊断有助于降低筛查门槛，在家庭或社区就可以实现大规模低龄婴幼儿群体的无感筛查，提前做好疾病预警和加速干预流程。本文首先回顾自闭症婴幼儿的传统筛查与诊断工具，然后对近 10 年自闭症婴幼儿(0~3 岁)智能化识别的研究进展进行分类梳理。其他年龄段的自闭症儿童和青少年的新型智能化识别研究也进行综述，这些研究所使用的数据采集手段和智能化识别技术对自闭症婴幼儿的智能化识别具有较高的启发和借鉴意义。最后，本文探讨了尚待解决的问题和未来的研究方向，为建立我国人工智能辅助的自闭症早期筛查与诊断体系提供新思路。

2 传统的自闭症筛查与诊断方法

自闭症最早的症状出现在生命的前一两年(Matson & Goldin, 2014), 50%的父母报告患儿在 2 岁时表现出症状, 90%的父母报告患儿在 3 岁时表现出明显的症状(Matson et al., 2008)，而诊断的年龄通常为 3 岁(Gilmore et al., 2018; Pierce et al., 2019)。自闭症患儿一旦错失最佳干预时期，后续治疗手段对康复效果将大打折扣。因此，很有必要尽早诊断和干预治疗。近年来，国内外学者均提倡自闭症的早期筛查，即对 18~24 个月的幼儿进行初级筛查，一旦发现可疑症状，即刻进入重点筛查。若筛查结果显示存在风险，则需转入早期诊断模式，尽早进行干预以获得最佳康复效果(Hyman et al., 2020)。可以说，早期筛查是早期诊断的基础，早期诊断是早期干预的前提。

传统的自闭症早期筛查方法主要分为基于量表形式的抚养者报告或专业观察和基于游戏任务的观察检查表两种形式。测评工具最早可适用于患儿 6 个月，且年龄跨度通常至少 6 个月(见表 1)。常用的一级筛查工具有幼儿自闭症筛查表(Checklist for Autism in Toddlers, CHAT)、CHAT 修改版(M-CHAT)、广泛性发展障碍筛查测验(Pervasive Developmental Disorder Screening Test, PDDST)，自闭症特质早期筛查表(the Early Screening for Autistic Traits, ESAT)等，适用于基层保健场所，大多由抚养者报告即可完成。其中, CHAT 是经过最严格的研究和验证、适用于婴幼儿的自闭症检测工具(尤娜, 杨广学, 2006)。常用的二级筛查工具有自闭症行为核查表(Autism Behavior Checklist, ABC)，儿童自闭症特质的婴幼儿筛查表(Baby and Infant Screen for Children with Autism Traits, BISCUIT)，2 岁儿童自闭症筛选测验(Screening Tool for Autism in Two-Year-Olds, STAT)，婴儿自闭症观察量表(the Autism Observation Scale for Infants, AOSI)等。二级筛查一般需要幼儿在场，由专业人员观察检测。

自闭症诊断的主要依据是美国精神学会(American Psychiatri Association, APA)精神障碍诊断和统计手册第五版(DSM-5)，常用的诊断工具是被称为“金标准”的自闭症谱系障碍诊断观察表第二版(Autism Diagnostic Observation Schedule, Second Edition, ADOS-2)和自闭症诊断访谈量表–修订版(Autism Diagnostic Interview-Revised, ADI-R) (Akshoomoff et al., 2006; Lord et al., 1994)。前者是在标准化活动情境下对婴幼儿直接进行观察，后者是对抚养者的半结构化访谈，两者均由经过培训的专科医生作出评估。

部分传统的自闭症早期筛查与诊断工具得到广泛认可，但在评估方式和使用效率上存在局限性，无法满足大规模的筛查和诊断需求，主要原因有：1)自闭症早期症状及风险信号需要依赖专科医生的评断，对观察者有专业性要求(Taylor et al., 2017)。评估者的专业水平、机构的医疗资源、文化背景差异均会影响自闭症评定结果的信度和效度(de Belen et al., 2020)。2)自闭症的确诊需要经历抚养者判断、医生访谈、临床观察和评估等，过程耗时且费用高昂(Wiggins et al., 2006)。3)自闭症症状表征非常广泛，部分临床症状在 2、3 岁前并无稳定表现(陈顺森 等, 2011)。此外，受环境、经济因素等限制，医生仅能基于有限的项目做出诊

表 1　ASD 早期筛查常用工具表

研究者	工具名称	缩写	适用范围(单位：月)	形式
Baron-Cohen 等(1992)	Checklist for Autism in Toddlers	CHAT	18~24	父母报告 专业观察
Robins 等(2001)	Modified Checklist for Autism in Toddlers	M-CHAT	16~30	父母报告 专业观察
Siegel (2004)	Pervasive Developmental Disorders Screening Test-II	PDDST- Ⅱ	12~48	父母报告
Dietz 等(2006)	the Early Screening for Autistic Traits	ESAT	14~15	父母报告 临床观察
Reznick 等(2007)	the First Year Inventory	FYI	9~12	父母报告
Krug 等(1980)	Autism Behavior Checklist	ABC	>18	父母报告
Schopler 等(2010)	Childhood Autism Rating Scale	CARS	>24	父母报告 专业观察
Matson 等(2009)	Baby and Infant Screen for Children with Autism Traits	BISCUIT	17~37	父母报告
Stone 等(2000)	Screening Tool for Autism in Two-Year-Olds	STAT	24~36	基于互动项目的观察
Bryson 等(2000)	the Autism Observation Scale for Infant	AOSI	6~18	基于半结构化游戏活动的观察
Young (2007)	Autism Detection in Early Childhood	ADEC	12~36	基于游戏项目的观察

断，无法对患儿在自然状态下的行为展开长期观察，在这种情况下，对症状的评估往往是不充分的。因此，目前亟待研究者提出新的技术手段，在保证结果达到一定准确度的前提下简化自闭症的筛查与诊断流程，缩减评估所需的时间与人力成本。

目前，人工智能技术辅助的自动化医疗诊断领域发展迅速，例如，依托计算机视觉的面部检测技术已实现超过 30 多种疾病的症状识别或预诊断，其中包括多种精神类疾病，如注意缺陷与多动障碍和抑郁(Thevenot et al., 2017)等。应用智能化手段识别自闭症婴幼儿的优势在于：1)可获取自然的、多维度的、多模态的行为数据进行综合分析，保证评估结果的有效性和客观性，提供可靠的预诊断信息，辅助医生的临床诊断；2)计算机视觉技术能捕捉肉眼无法观察和量化的自闭症婴幼儿的细微动作，有效识别异常行为或发现新的自闭症早期风险标志，相比人工筛查和诊断的成本和侵入性低，可应用于家庭或社区医院环境。

3　自闭症早期患者的智能化识别技术

尽管研究者已发现诸多自闭症的核心症状和早期风险标志，但低龄幼儿很难确诊，主要在于自闭症谱系障碍在不同亚型、具体症状和严重程度上表现出较大差异。并且，自闭症婴幼儿的行为表现通常伴随早期发育特征(Vyas et al., 2019)，同时依赖非自闭症的影响因素，例如认知功能和年龄等(Li et al., 2019)。然而，依靠抚养者报告早期症状易存在回忆偏差。临床医生的观察有限，需要患儿予以配合，存在取样偏差。如果能汇总自闭症婴幼儿的大量行为，特别是自然状态下的日常行为，使用更为客观的方法综合患儿的多方信息作出评估，将极大提高筛查与诊断的准确性和可靠性。

计算机视觉、智能传感器、机器学习、深度学习等人工智能技术已逐渐成功用于自闭症的早期预警(Hazlett et al., 2017)和机器辅助治疗(Zheng et al., 2015)。同时，自闭症的诊断和治疗领域每天都产生大量数据，基础数据的积累量已达到一定规模，合理使用历史数据可有效提高自闭症患儿的诊疗效率(廖梦怡 等, 2021)。本文根据系统评价和元分析(PRISMA)指南(Moher et al., 2009)，分别对 Web of Science、PubMed、IEEE Xplore、ProQuest 数据库中发表于 2010~2020 年间的文献以关键字——“自闭症谱系障碍” (“自闭症”)和“机

器学习” (“深度学习”, “计算机视觉”, “情感计算”)为主题词进行搜索, 去除重复项后初步检索到741 篇文献。结合“自闭症早期患者的智能化识别”的研究主题, 对所有文献按固定标准进行筛选, 包括: 1)关注人类自闭症, 剔除动物类研究; 2)主题集中在智能化技术在自闭症筛查、诊断中的应用, 而非以干预、治疗为主的情境; 3)剔除含有基因、生物标记物等生物、医学类的研究; 4)对象以婴幼儿、儿童、青少年患者为主, 自闭症成年及老年患者非特殊原因不予考虑; 5)研究目的是自闭症检测或自闭症经典测评任务中的风险行为, 而非衍生行为(如, 自闭症患者的自伤行为、睡眠等)。最终获得 576 篇目标文献, 见图 1。从上述文献发现, 10 年间自闭症自动化识别的探索基于不同类型的数据展开, 可依此归纳为 6 个研究子领域: 1)基于经典任务行为数据的识别(114 篇); 2)基于面部表情和情绪数据的识别(144 篇); 3)基于眼动数据的识别(18 篇); 4)基于脑影像数据的识别(169 篇); 5)基于运动控制和运动模式数据的识别(58 篇); 6)基于多模态数据的识别(73 篇)。本文根据文献引用率排序法和滚雪球法选择重点文献并进行文献扩充, 最终结合 80 篇重点文献进行综述, 下文将从这 6 个领域逐一展开介绍。

3.1 基于经典任务行为数据的自动识别

自闭症早期诊断的第一步就是对早期潜在危险信号的筛选、评估和处理(陈顺森等, 2011)。长期以来, 自闭症的早期异常行为在回顾性研究(父母报告、家庭录像分析)和前瞻性研究、早期筛查量表、临床诊断中得到充分研究和验证, 积累了较多经典的临床评估任务及相应的行为观测指标, 如 AOSI 中的叫名反应任务、视觉追踪任务、注意脱离任务等。近年来, 研究者针对自闭症早期经典任务中的异常行为提出自动化检测模型。研究者一般采用非接触式视觉系统及传感器技术(电子设备前置摄像, RGB 摄像机, Kinect 3D 体感摄影机等)采集患儿在任务中的面部表情变化、头部运动、肢体运动、声学等多维度行为数据, 提出基于任务的异常行为检测算法和自动化评估模型, 以替代传统的人工观察与评估, 提高筛查效率。下文以叫名反应任务、视觉注意力任务为例, 介绍行为数据采集技术, 行为数据采集过程、异常行为检测算法与预测模型。

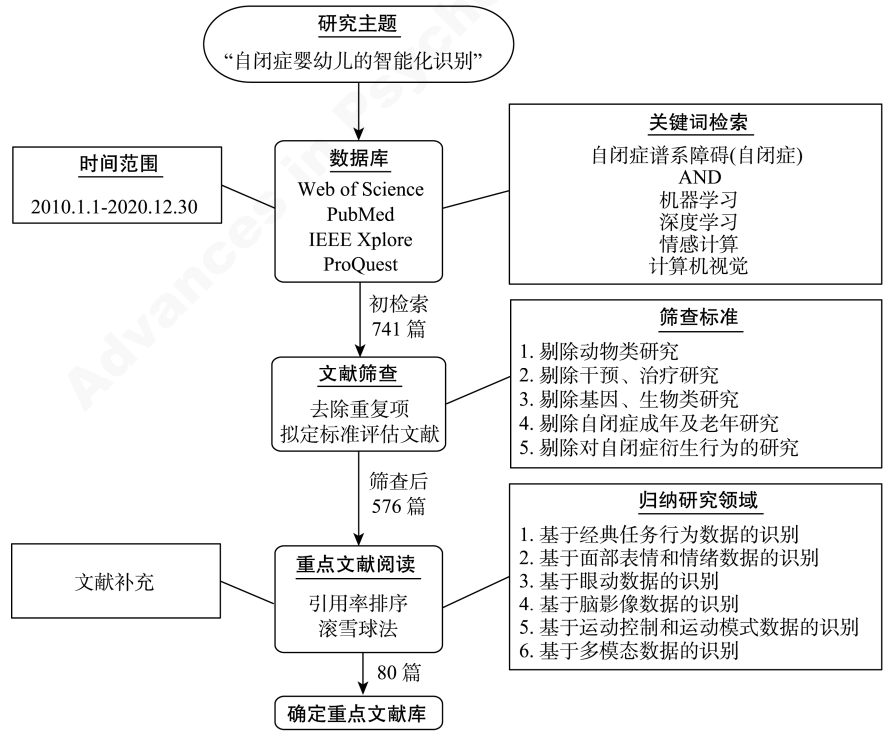


图 1 文献调研流程图

“叫名反应” (Response to name, RTN)是自闭症早期筛查量表和临床诊断中最常出现的经典任务之一。婴儿自 4~6 个月开始就对自己的名字有反应，在听到自己的名字时会有选择地转动头部，表明他们已领会到名字有打招呼的含义(Imafuku et al., 2014)。叫名反应需要专业人员现场观察记录或事后注释，根据计分手册或诊断标准进行评估。事实上，自闭症患儿在叫名反应中的非典型(atypical)行为可量化为可计算的观测指标，如患儿的眼睛注视、头部姿态变化、名字呼叫后的反应时长等等。例如, Bidwell (2014)等人分析了公开标注的多模态二元行为数据集(Multimodal Dyadic Behavior Dataset, MMDB)中的 50 组 15~30 月龄的幼儿参与“叫名反应”的音视频，其通过天花板高架式 Kinect、前置摄像机与追踪器估计幼儿头部姿态变化。研究中，幼儿的头部偏航(yaw)夹角和听到姓名后的反应潜伏时长作为预测幼儿对社交刺激(呼唤姓名)积极(消极)反应的行为指标，不同分类器的预测效果略有差别，最高精确率(precision)和召回率(recall)达到 89.4%和 83.3%。Wang 等人(2019)对“叫名反应”任务构建了包含实验流程、数据采集和自动评估的自闭症辅助筛查系统，有助于缩减筛查的人力成本，有望在医疗欠发达地区发挥作用。研究者在实验环境搭建了由 Kinect 和 2 个 RGB 摄像机组成的多传感器系统，可同时对幼儿(平均年龄 2 岁)的面部、注视、姿态和语音等行为信息进行采集和集成，实现行人检测和骨骼点提取(Microsoft Kinect SDK)、面部表情识别(Baltrušaitis et al., 2015)、面部标记点的检测与追踪(Baltrušaitis et al., 2013)、眼球中心定位(Wang et al., 2018)和头部姿态估计(Baltrušaitis et al., 2016)等。研究者基于眼球中心定位和头部姿态估计算法，使用人眼注视方向的旋转角度和注视持续时间作为预测幼儿积极(消极)反应的行为指标，平均分类准确率(accuracy)为 92.7%。

非典型注意力的评估也是最常见于自闭症早期筛查与诊断工具的经典任务之一。目前，基于自闭症幼儿的视音频、图像等能够自动识别多种非典型的注意力特征，例如，不流畅的视觉追踪(Zwaigenbaum et al., 2005)、人脸面部的注视频次低(Ozonoff et al., 2010)、注意力分离能力弱(Elsabbagh et al., 2013)等。Hashemi 等人(2014)使用面部检测与追踪技术对 AOSI 中两种非典型视觉注意力任务进行自动评估，分别是：1)注意脱离：从两个竞争性视觉刺激中分离注意力并移动视线; 2)视觉追踪：视觉跟踪移动的物体横向穿过中线。研究者使用 GoPro Hero 运动摄像机记录了 12 名 5~18 月龄的自闭症风险婴儿的多个实验试次，基于偏航(yaw)和俯仰(pitch)的头部姿态运动评估视觉注意。结果表明，自动评估与专家的评分者信度(Cohen’s Kappa)为 0.75，远高于非专家评分(0.27~0.37)。Bovery 等人(2019)开发了移动设备端的实验任务测量自闭症幼儿的非典型注意力，左右屏分别呈现社交与非社交的影像刺激，前置摄像记录 104 名 16~31 月龄幼儿观看刺激时的面部动态影像。研究者通过计算 51 个面部标记点(Hashemi et al., 2015)与 3D 标准人脸模型(3D canonical face model)间的旋转参数估计头部姿态(Fischler & Bolles, 1981)，结合头部偏航(yaw)夹角和眼球虹膜位置估计注意方向，进而测量自闭症幼儿对不同刺激的注意时间、注意偏好和注意转移。Campbell 等(2019)采用上述类似的实验范式对 16~31 月龄自闭症幼儿的非典型注意和“叫名反应”进行自动评估。结果表明，叫名反应的自动评估与专家评分具有较高一致性(ICC = 0.84, 95% CI 0.67~0.91)，敏感性(sensitivity)为 96%，特异性(specificity)为 38%。

现有研究集合了多种自闭症的经典临床评估任务，开发了集任务、数据采集、算法于一体的移动端应用程序，形成集成化、低成本、可扩展的自闭症自动筛查工具，应用范围从实验室研究扩展至初级保健院、学校、家庭等社会医疗场所，已实现自闭症幼儿的非典型情绪、社会参照、社会性微笑、叫名反应等社会互动行为的自动检测，已具有一定的预测效果(Hashemi et al., 2015; Hashemi et al., 2018)。然而，目前仅涉及较为简单的评估任务，尚未涉及复杂任务，因为幼儿在这类任务中的反应模式更加多元，大大增添了自动化检测的难度。例如, ADOS 中的“泡泡游戏”是为评估幼儿和他人“共享乐趣”的能力，需要首先建立“共享乐趣”对应的行为维度体系，只有同时结合幼儿的表情、眼神、自发动作、声音等多模态时序数据进行协同建模才有可能达到较好的检测效果。

3.2 基于面部表情和情绪数据的自动识别

社交沟通障碍是自闭症患儿区别于典型发展群体的重要特征，表现在社会情感互动和非言语

沟通两个方面，如受损的面部表情模仿能力、面部表情多样性及程度等，是临床常用的面部行为指标。使用面部表情分析技术能够克服人类感知的局限性，快速、客观地自动识别自闭症。

近年来，计算机视觉技术的进步推动了人工智能情绪识别的发展，其主要任务是基于面部图像或视频开发算法识别人脸情绪标签，如基本情绪分类(de Belen et al., 2020)。研究者尝试构建用来检测异常情绪认知和表达的算法，并将其应用到自闭症婴幼儿的自动化识别中。然而，受限于自闭症患者取样的特殊性，样本规模普遍偏小，但正常人群的情绪识别领域积累了较多模型和公开数据集，因此，一种研究思路是先基于正常人群数据集的面部特征迁移或调整已有模型。例如，Han 等人(2018)基于 FERET 和 Cohn-Kanade (CK+)正常人脸数据集，抽取并比较了正常人群和自闭症儿童的面部表情特征差异，提出了一种基于稀疏编码(Sparse coding)的特征迁移学习算法。在自闭症儿童与机器人互动的过程中，实时采集面部数据并识别情绪类型，平均准确率(accuracy)在80%以上。

研究者在实验室或自然环境下采集婴幼儿静态面部图像或动态面部视频，构建自闭症婴幼儿的情绪识别或自闭症分类模型，已达到较为精准的预测效果。例如，社会性微笑是自闭症早期的重要风险标志(毕小彬 等，2020)，特别是母婴互动中婴儿的微笑是检测自闭症的关键信号，在临床和家庭环境中自动识别婴儿微笑有助于提升早期筛查效率。例如，Tang 等人(2018)基于母婴互动时 34 名 6~24 月龄婴儿(自闭症高风险 11 人)的面部视频数据集(RCLA&NBH_Smile，含 77000 人工标注视频帧)训练卷积神经网络(Convolutional Neural Networks，CNN)自动检测视频中的婴儿笑容，平均准确率(accuracy)为 87.16%。Li 等人(2019)发现面部表情、面部动作单元、情绪唤醒度和情绪效价是自闭症分类的重要面部特征。研究者通过移动设备前置摄像记录 105 名儿童观看视频时的面部动态影像，使用基于公开数据集 AffectNet 和 EmotioNet 预训练的 CNN 模型和时序特征抽取方法获得面部特征表示，建立二分类预测模型，自闭症分类的敏感性(sensitivity)和特异性(specificity)为 0.76 和 0.69。Shukla 等人(2017)提出从面部图像自动检测发育障碍的方法，包括自闭症谱系障碍、脑瘫、胎儿酒精综合症、唐氏综合症、智力障碍、早衰等等。研究者使用微调的 CNN 模型——Alex Net 对两千余张人脸图像获得面部特征表示，使用支持向量机(Support Vector Machine, SVM)构建不同疾病的二分类任务。结果表明，自闭症分类的平均精确率(precision)为 93.33%，且模型对不同疾病的分类优于人工分类(非专家)的结果。

在对自闭症患者面部情绪的分类基础上，另有研究深入探讨了自闭症患者面部表情的发生过程。研究者通过面部动作单元编码等方法检测、追踪人眼无法观察到的微观面部运动特征，对自闭症患者产生面部表情的能力进行量化评估，如检测个体是否开始产生表情、产生特定表情时面部各区域和肌肉群的激活情况，这有利于自闭症的精细诊断和对点干预，对自闭症婴幼儿的智能化筛查及其面部表情特征研究具有启发性。Leo 等人(2019)提出对自闭症患者表达面部情绪的能力进行计算分析的方法，构建了自动化评估框架，包括 4 个算法模块：人脸检测、人脸特征点检测与追踪、面部动作单元强度估计、面部表情分析。17 名自闭症患者(6~13 岁)和 10 名典型发展幼儿(26~35 月龄)产生 4 种基本情绪的面部动态影像作为分析材料。结果表明，该方法能够分别较为准确地预测上述两类群体表达特定情绪时的专家评定分数(二分类任务)，自闭症患者情绪识别的精确率(precision)和召回率(recall)为 0.90 和 0.85。此外，结果发现自闭症患者在表达快乐、恐惧和愤怒情绪时能同时使用上面部和下面部，而表达悲伤情绪时的表情主要集中在下面部。Guha 等人(2016)使用运动捕捉技术(motion capture technique, mocap)对 9~14 岁的高功能自闭症(high functioning autism)患者面部表情的细微动态特征进行研究，要求被试模仿固定序列的面部情绪，同时使用 6 个红外运动捕捉相机以每秒 100 帧的速度记录 32 个面部标记点的运动。多尺度熵方法(multiple scale entropy, MSE)分析个体面部动态复杂性发现，高功能性自闭症患者表达情绪时面部表情的动态复杂性较低，缺乏丰富的面部表情模式和变异性，差异主要集中在眼部区域。Ahmed 和 Goodwin (2017)则着眼于面部表情计算分析在自闭症患者的计算机辅助教学情境中的应用，研究如何通过面部表情变化测量学习投入水平，以达到辅助教学的目的。研究者基于自闭症青少年

(平均年龄约 12 岁)学习过程中计算机前置摄像记录的面部反应视频，使用面部动作编码系统(Facial Action Coding System, FACS)对特定情绪引发的面部动作单元进行编码，以计算机表情识别工具(Computer Expression Recognition Toolbox, CERT)得到视频中个体头部朝向屏幕的时间比和面部动作单元的激活状态作为监控学习投入的行为和情绪指标。

自闭症患者的面部情绪识别和面部表情特征一直是自闭症研究的热点。自闭症面部表情和情绪的自动化识别能够解决面部人工编码耗费时间长，难以分析大样本、实时分析的局限。目前，该领域的研究可根据表情产生的类型(自发表情，模仿表情)、情绪诱发刺激类型(视频，感官刺激，社交互动)、数据(静态面部图像，动态面部视频)、面部表情自动化评估的目标(定性，定量)划分为不同的子领域。同时，研究成果可用于开发“治疗机器人”，实时、自动识别自闭症幼儿的情绪并进行对点干预，辅助临床医生“阅读”自闭症儿童的面部表情，提高诊疗的有效性。最后，建立面部表情和情绪识别算法的第一步通常是对采集的大量面部视频帧进行标注，而完全采用人工标注费时费力，从众包平台(crowdsourcing platform)的外包标注又往往会存在评分者一致性低下的问题。Kalantarian 等人(2019)提出了 3 种自动标记算法对儿童(平均年龄为 8.5 岁)面部表情视频帧进行 6 种基本情绪的自动标注，包括：厌恶、中立、惊奇、害怕、愤怒和快乐，结果表明前 4 种情绪的自动标记效果相对较好。

3.3 基于眼动数据的自动识别

眼神交流是非言语沟通的关键要素，表示个体对社交互动的兴趣、关注和参与，是识别语言障碍、情感状态及自闭症早期风险标志的重要指标。目前，已有充分证据表明自闭症幼儿与典型发展群体的注视方式存在显著差异，如非典型的凝视、眼神交流和共同注意(Chong et al., 2017); 对社交和非社交图像的偏好不同(Campbell et al., 2014; Chawarska et al., 2013; Shi et al., 2015)。

眼动追踪是测量社会性知觉和社会性偏好的常用方法之一，其能够捕获眼神的运动轨迹，非常适合对感知异常的自闭症患儿进行研究。传统的眼动追踪主要有两种方式，其一是佩戴头戴式眼动设备，低龄幼儿需要较长时间适应；其二是视点追踪(viewpoint tracking)，能追踪的视线范围限制于屏幕，仅适用于高度受控的实验室研究，无法测量自闭症患儿在社交环境中的注视行为(Chong et al., 2017)。因此，研究者开始探索使用非接触性的眼动追踪技术，如基于图像中的人眼外观(Lu et al., 2014)或构建人眼数学模型(Li & Li, 2015)，同时对眼动数据中蕴含的心理因素进行分析。例如，Syeda 等人(2017)研究了自闭症患者(5~17 岁)在处理人脸情绪图像时的面部扫描模式和情绪识别能力。研究者使用安装于笔记本的眼动仪(Tobii EyeX Controller)采集自闭症患者和典型发展个体在观看 6 种基本情绪的人脸图像时的眼动数据。研究发现，自闭症患者在扫描面部时较少关注面部核心特征(眼睛、鼻子、嘴巴)，因而很难正确感知他人情绪。Chrysouli 等人(2018)基于正常人群和自闭症患者的眼睛凝视影像数据集(MaTHiSiS)识别学习者在人机交互学习情境中的情绪状态，通过构建两阶段的双流 CNN 模型(two-stream CNN)，融合眼部连续图像帧间的光流信息(optical flow)和静态图像帧的空间信息识别个体当前是否处于投入、无聊或沮丧的情绪状态。近年来，研究者评估自闭症儿童与抚养者面对面互动状态下的眼神接触时通常采用 POV (point-of-view)相机采集眼部数据，即要求成年人佩戴头戴式摄像头记录儿童的注视行为。例如, Chong 等人(2017)基于 POV 相机采集的数据，提出检测自闭症儿童在与抚养人自然互动过程中的脸部朝向和眼神接触的算法。他们基于 100 名自闭症儿童(3~6 岁)和典型发展幼儿(18~36 月龄)的、包含 156 个互动片段、共 22 小时的影像数据集，开发了一种端到端的深度学习框架(Pose-Implicit CNN)检测儿童的眼神接触，结果优于其它模型(AlexNet, PEEC, GazeLocking)，精确率(precision)为 0.78, 召回率(recall)为 0.80。

传统方法收集自闭症低龄幼儿的眼动数据几乎不可能在大规模筛查中施行，其一是需在高度受控的实验室环境中借助专业仪器测量；其二要求被试长时间持续注视屏幕，不适合对低龄幼儿进行测试，因而研究者很难在自然社交互动中(如, 幼儿与抚养者)评估眼神接触或注视。而上述非侵入性的眼动测量技术将有利于自闭症早期眼动特征的相关研究，如通过摄像机记录幼儿的面部(主要是眼睛)和头部姿态变化，据此分析幼儿注视的

位置和凝视时长。需注意的是，不同于实验室的标准化环境，幼儿在家庭或保健院等场所中采集的视频数据可能存在面部遮挡、头部位置偏移的问题，需进行视频预处理和校正等。

3.4 基于脑影像数据的自动识别

自闭症的精准诊断对自闭症患者的及早干预和及时治疗至关重要。目前国际上开展了大量针对自闭症精准诊断的研究，力图找出自闭症的行为学、遗传学和影像学标志物(Hong et al., 2020; Lord et al., 2020; Talbott & Miller, 2020; Wolfers et al., 2019)，同时结合人工智能技术，实现对自闭症的客观诊断。然而，这些研究绝大多数是针对儿童和成人的(Dickinson et al., 2021)，针对婴幼儿的自闭症智能诊断研究尚少。

脑影像技术的发展大大推进了人类对于自闭症病理机制的理解，脑影像技术与人工智能技术的结合为自闭症的早期精准诊断提供了新的契机。当前自闭症诊断难、诊断准确率不高的主要原因在于自闭症本身的病理机制存在异质性，而脑影像技术在获取脑结构及功能精细信息进而捕捉不同病理亚型的特异性特征方面具有很大优势(Emerson et al., 2017)。因而，基于脑影像的自闭症客观诊断研究受到极大关注。目前广泛应用于自闭症精细化诊断的影像技术主要包括脑电(EEG)、结构磁共振成像(sMRI)和功能磁共振成像(fMRI)。

结构磁共振成像(sMRI)能够捕捉自闭症婴幼儿细微的脑结构变异，因而在自闭症的早期诊断中有良好表现。Hazlett 等人(2017)在 Nature 发表文章，报告了他们基于 sMRI 进行自闭症早期诊断的研究：基于6~12月龄的148个婴幼儿样本的sMRI数据，他们提取了皮层厚度、皮层表面积和脑体积等特征，结合深度学习算法构建了自闭症早期诊断模型，最终达到了 81%的敏感性(sensitivity)和 88%的特异性(specificity)。脑电(EEG)的高时间分辨率便于其精准反映自闭症婴幼儿的脑功能时空共变模式的异常，可为自闭症的早期诊断提供有效特征。Gabard-Durnam 等人(2019)基于 171 例 3~36 月龄的婴幼儿的纵向 EEG 数据，结合 Logistic 回归对自闭症患者和正常个体进行分类(类别标签为样本 36 月龄时的诊断结果)，研究发现出生后第一年的 EEG 功率动态波动(相较第二、三年)对于自闭症早期诊断最为有效(正确率达 91%)。Dickinson 等人(2021)基于 65 例 3 月龄样本的 EEG 数据，结合支持向量回归算法，对样本在 18 月龄时测得的自闭症行为评分进行预测，预测值和真实值的相关系数高达 0.76。功能磁共振成像(fMRI)同时具备高的时间分辨率(相对于正电子发射断层成像(PET)、单光子发射计算机断层成像(SPECT))和空间分辨率(相对于 EEG)，能够提供极为丰富的脑功能活动及脑功能网络的静态和动态特性信息，为自闭症的客观诊断提供有效特征。Emerson 等人(2017)基于 59 例 6 月龄样本的静息状态 fMRI 数据，提取功能连接特征，以样本在 24 月龄是否被诊断为自闭症作为标签，用支持向量机构建分类模型，诊断敏感性(sensitivity)为 81.8%，特异性(specificity)高达 100%。

借助脑影像数据有望实现针对婴幼儿的自闭症精准诊断，但出于婴幼儿样本数据的获取较儿童及成人样本更加困难，以及研究人员学术背景的限制(自闭症的智能诊断需要医学、神经影像学以及计算机科学三个学科领域人员的通力协作)，目前国际上针对婴幼儿的自闭症智能诊断的研究尚处于起步阶段，具体表现在：第一，已有研究主要致力于从正常对照样本中区分出自闭症，还没有对自闭症进行更为精细的分级和分型研究。大量研究表明自闭症是一种高度异质性疾病(Elsabbagh et al., 2013)，在对婴幼儿是否患有自闭症作出定性判断之外，如果能提供自闭症严重程度的判断、所属的病理亚型等信息，将大大有助于为患者制定更为适宜的个性化治疗方案。第二，多数研究仅提取了较为粗略的脑影像特征，未充分发掘影像数据中包含的有效信息。人脑是自然界最为复杂的系统，自闭症患者的脑功能异常必然表现为复杂时空共变信息的异常。目前针对自闭症诊断的研究都还是基于静态功能连接、局部一致性等较为简单的特征开展的，近期研究表明自然状态 fMRI 更适于婴幼儿脑功能研究(Xie et al., 2021)，而脑功能网络的动态属性能提供比静态功能连接更为丰富和有效的脑功能信息(Eslami et al., 2021)。第三，当前研究多采用经典的机器学习算法，还未充分利用性能优越的深度学习算法。目前只有较少的研究采用深度学习技术来建模，例如，Xu 等人(2020)基于功能性近红外光谱(fNIRS)时序数据构建长短期记忆模型(LSTM)和卷积神经网络(CNN)结合的深度学习模型对自闭症儿童和典型发展儿童(平均年龄约 9 岁)

进行分类预测，结果表明，模型表现出高度的分类准确性，敏感性(sensitivity)和特异性(specificity)分别为 97.1%和 94.3%，此外，准确率相比原有模型提升 8%，证明使用深度学习模型能够有效提升自闭症患儿识别的准确性。

3.5　基于运动控制和运动模式数据的自动识别

非典型的运动控制和运动模式也是自闭症的早期特征。Landa 等人(2006)发现，自闭症患儿早在 14 个月和 24 个月时在 MSEL 测验(Mullen Scale of Early Learning)中的精细动作和粗大动作得分较低。此外，许多研究报告了自闭症患儿在俯卧、仰卧、爬行、走路上存在姿势异常、动作不协调、运动控制能力弱的现象(Esposito et al., 2009; Teitelbaum et al., 1998)。这些发现提供了基于婴幼儿时期的非典型运动模式识别和预测自闭症的支持性证据。传统方法对自闭症患儿的运动功能的评价通常基于父母报告和专家现场观察，编码方法和评价标准通常针对某个具体的研究场景，缺乏效度验证(Ozonoff et al., 2008)。现有研究大多基于自闭症儿童的运动视频构建自闭症早期运动控制和运动模式的自动检测方法。例如，Dawson 等人(2018)基于视频的面部自动化检测技术评估自闭症幼儿与典型发展幼儿在自发性注意状态下的头部姿势控制能力。研究者在实验室采集了 106 名 16~31 月龄的自闭症幼儿和典型发展幼儿观看动态泡泡、机械兔子等视频时的面部动态影像，通过识别、追踪、计算相邻视频帧中既定面部标记点的坐标位移对头部运动进行量化。结果表明，自闭症幼儿的头部运动速率显著高于典型发展群体，即自闭症幼儿在注意状态下难以保持头部的中线位置。Martin 等人(2018)采用相同实验范式采集了 2.5~6.5 岁儿童观看社交和非社交刺激视频时的上半身视频录像，不同于计算面部标记点位移的方法，该研究使用 Zface (http://zface.org/)算法(Jeni et al., 2015)，该算法被证明有高计算效率和精度(Jeni & Cohn, 2016)，可逐个对 2D 视频帧实时形成高密度的 3D 面部形状，进行三维尺度上的头部运动追踪，x、y、z 轴分别对应头部俯仰(pitch)、偏航(yaw)、侧倾(roll)。通过计算连续视频序列各坐标轴的角位移和角速度来量化头部运动强度，研究发现自闭症儿童在观看社交刺激时，头部运动水平和移速均高于典型发展儿童，疑似自闭症儿童通过头部运动调节对社交刺激的感知。

许多研究还尝试基于视频序列中的运动特征建立自闭症儿童的自动识别模型。Zunino 等人(2018)关注自闭症儿童(平均年龄约 9.8 岁)的抓握行为，分析对象包含儿童抓握、放置、传递水瓶的视频动作序列数据集，平均长度 83 帧。研究者使用卷积神经网络(CNN)和长短期记忆网络(Long short-Term Memory, LSTM)构建自闭症分类模型，不仅能判断视频帧中的行为影像是否为自闭症个体，同时基于 LSTM 的隐藏层的时空表征信息和内部门机制，还可输出可视化的注意力地图(normalized attention map)，即在图像帧上表示出对自闭症分类有作用的、提供重要判别线索的位置，具备可解释性，为临床工作者进一步提供支持性的辅助信息。Vyas 等人(2019)使用自闭症远程诊断服务 NODA 项目(https://behaviorimaging.com/)的数据，包含家长记录的儿童日常活动视频 555 个，均已由专家标记出自闭症诊断结果。在运动序列特征的表征方面，采用预训练的 2D Mask R-CNN 深度学习网络(Girdhar et al., 2018)识别图像帧的 15 个人体部位关键点执行姿态估计，并应用粒子滤波器(Particle filter)对判断缺失的关键点进行插补(Arulampalam et al., 2002); 使用 Pose Motion (PoTion Representation)表征为身体关键点随时间变化的 RGB 热图(Choutas et al., 2018)，并输入最终的 CNN 分类器中。结果显示，分类的准确率(accuracy)、精确率(precision)和召回率(recall)分别是 72.4%、72%和 92%。该研究通过关键点轨迹刻画身体姿势的变化，使用具有解释性的浅层行为信息且模型可输出身体关键点变化热图，帮助研究者直观理解自闭症儿童的运动特点。

随着硬件设备技术的快速发展，惯性运动传感器、陀螺仪和磁力计已经被集成到智能手机、平板电脑和智能穿戴设备中，用于采集个体的运动数据。Anzulewicz 等人(2016)探索了严肃游戏场景下检测自闭症及其运动模式的方法。研究招募了 37 名 3~6 岁的自闭症儿童和 45 名典型发展儿童，儿童手动操作平板游戏过程中，通过设备的触摸屏和内置惯性传感器(三轴加速度计、陀螺仪、磁力计)记录手部运动数据。研究者基于原始传感器数据获得的共 262 个特征构建多种机器学习自闭症分类模型，十折交叉验证的结果显示，正则化贪心森林算法(Regularized Greedy Forest, RGF)达到最佳分类效果，敏感性(sensitivity)和特

异性(specificity)为 83%和 85%。此外，研究发现自闭症儿童在玩游戏时的手部撞击力、手势压力、用力的分配、屏幕点击速率等与典型发展儿童存在显著差异。

目前针对自闭症的多种非典型运动模式的自动检测研究已有较大进展，从涉及粗大的姿态、肢体动作到精细化的头部、手部动作，包括：身体运动的整体姿态变化、手部运动(抓握、放置、传递物品)、注意状态下的头部运动(头部运动速率、平衡性)等。人体运动数据的采集对硬件设备具有较高要求，随着智能传感器技术的不断发展，研究者可采用智能穿戴设备和体感设备采集并记录个体的运动数据，对低龄幼儿和儿童的非典型动作进行自动化识别，以此对自闭症早期运动发展规律进行深入研究。

3.6 基于多模态数据的自动识别

自闭症儿童的心理、生理和认知状态可同时通过面部表情、身体姿态、眼睛注视、语音、文字以及生理信号等多种维度信息进行反映。并且，由于实际测量环境的异质性和临床样本的稀缺性，单模态数据往往无法携带足够的信息用于识别(Chen & Zhao, 2019)。目前，自动化识别的趋势是采用多模态数据融合思路，对多种模态数据的相关特性或中间决策进行融合分析，获得更有价值的数据或高层信息以提高预测的准确性(Poria et al., 2017)，性能优于单模态建模(de Belen et al., 2020)。例如，Chen 和 Zhao (2019)分别采用拍摄任务 (Photo-Taking task) 和图像观察任务 (Image-Viewing task)，基于自闭症患者异常的注意偏好建立自闭症自动化识别模型。在建模阶段，研究者借鉴跨模态检索(cross-modal retrieval)思想，将两种任务的模态数据(眼动，图像)融合，建立共同的预测模型实现多模态信息的特征表征和信息互补。结果表明，在两种任务上，多模态信息建模方法的预测性能相较于传统单模态建模均有提升，拍摄任务的预测正确率从 76%提高至 84%，图像观察任务的正确率从 97%提高至 99%。廖梦怡等人(2021)构建了融合多模态数据的 3~6 岁自闭症儿童智能化识别方法。研究者使用眼动数据、面部表情数据、认知得分数据以及认知反应时数据，利用数据差异性分析进行特征选择，根据数据来源和时间同步性将数据进行分层融合。结果表明，与单模态识别方法相比，多模态建模方法与基于《自闭症行为评定量表》(Autism Behavior Checklist)的评估结果达到最高一致性。

另有研究基于自闭症儿童与社交机器人交互过程中产生的多模态行为特征进行自闭症自动化诊断和干预效果评估。例如，Scassellati (2007)界定了人机交互数据中反映个体社交技能的行为指标，以标准化的测量方式提升人工记录和评估的信度。具体包括三方面的行为指标：(1)凝视方向和注意焦点；(2)个体与他人的距离和位置追踪；(3)个体声音的韵律和语调。此外，研究者基于多模态数据中表征社交互动的行为指标，如面部朝向、相对位置、物理距离等行为数据，评价自闭症儿童(平均年龄 3.4 岁)与社交机器人互动的积极性和参与度(Feil-Seifer & Matarić, 2010; Moghadas & Moradi, 2018)。另外，目前有监测并提供自闭症患者社交技能服务的网络平台，通过非侵入式传感器和可穿戴设备采集患者与他人的多模态日常互动行为数据，如个体与交谈者的物理距离、姿势、上半身运动、面部微表情等社会计量器(sociometer)的行为指标，实时传输至Microsoft Azure云计算平台存储、分析和完成行为标注，产生可视化的数据分析报告并发送给治疗师，可有针对性地设计干预方案(Winoto et al., 2016)。

目前已积累较多公开的、可供研究者训练和微调模型的自闭症幼儿多模态数据集。例如，研究自闭症婴幼儿社交和沟通行为的多模态二元行为数据集(Multimodal Dyadic Behavior, MMDB)，包括 160 余组、平均 3~5 分钟的 121 名婴幼儿(15~30 月龄)与成年主试的半结构化互动片段。数据集已有行为评价框架和人工标定，即依照自闭症编码手册和评分细则对互动过程中的关键行为进行编码评分，如婴儿的注意力和眼神接触、社会性微笑、发声和言语表达(如，字、词组)、沟通姿势(如，指向、挥手、鼓掌)等(Rehg et al., 2013)。此外，还有用于辅助自闭症儿童干预治疗的多模态数据集，例如 DE-ENIGMA 大型公开数据集，包含 128 名自闭症患儿与治疗师、治疗机器人的长时段互动视频(13TB)，其中专家对 50 名患儿的数据标注了情绪效价、情绪唤醒、身体姿势等特征，可用于训练人工智能模型进行自动识别。

目前，大多数研究主要基于独立收集的临床数据构建检测模型，而针对于同一任务下的同类方法或模型未能得到比较，研究成果间成孤立之

势，因而需要借助公开数据集作为模型性能评价的“基准”。同时，越来越多的自闭症公开数据集提供了推进机器学习和深度学习发展所需的数据规模，研究者可使用公开数据集预训练模型，或使用不同模态的数据提升模型性能，增强检测方法的泛化性。

4　小结和研究展望

我国人口基数大，罹患自闭症的患者数量多，自闭症的诊断过程耗时，且费用高，需要长周期的特殊教育训练和行为干预。自闭症在幼儿期越早发现，越早干预，康复效果越好，治愈率越高。在 3 岁内开始治疗是最佳时机，随着年龄增长，治疗难度增大。然而，目前我国自闭症筛查与诊断处于“三缺”状态，即缺乏诊断标准、专业人员、康复路径，自闭症康复干预系统也不健全、整体水平偏低，治疗效果有限。自闭症的诊断主要依据医生的经验积累，缺乏便捷和客观有效的诊断工具，导致从父母发现儿童异常再到确诊需要漫长的等待时间，因而很多患儿会错过最佳干预期。此外，受限于诊断滞后，以及婴幼儿的行为语言能力尚待发育造成的干预困难，现有干预方法的适用对象大多是 3 岁以上的儿童和青少年，且每种干预方法的疗效并不明确，每种疗法下的干预手段比较单一，这也是制约患儿有效康复的主要原因。

因此，目前有必要革新现有的诊断流程和康复路线，完善智能辅助的自闭症早期筛查、诊断和治疗体系，缓解执行的时间压力和压缩人力成本。特别是对患者家庭，越早发现和干预，其预后效果越好，能够促进自闭症患者的发展进程、改善其语言、减少问题行为(陈顺森 等, 2011)，产生持久甚至终生受益的良好疗效，将家庭的经济和精神负荷降到最低。近年来，计算机视觉、语音技术、智能穿戴设备、脑成像等技术能够实现多种行为数据的采集和监测，方便研究者获取多模态、不同场景和多种行为维度的患者数据。尤其基于自闭症患儿在自然场景下与抚养人互动的视音频数据建模，相比传统方法，能够更加真实、全面地反映行为特征与症状间的关系，捕捉肉眼无法观察到的细微表情或肢体动作的变化。围绕建立婴幼儿自闭症的智能化无感筛查、诊断评估体系议题，本文提出当前主要存在的两点问题，未来研究方向如下所述。

4.1　缺乏针对婴幼儿的有效的筛查工具

目前，针对自闭症的早期筛查与诊断取得了很多进展，一些测量工具得到广泛认可，但同时自闭症的早期筛查与诊断也存在诸多不足之处，包括早期筛查难和普及难、检测评估不精准等。具体表现在以下三个方面。

第一，缺乏精细化的婴幼儿自闭症早期行为诊断体系。自闭症早期评估工具的年龄使用范围跨度较大，一般至少为 6 个月，且评估项目对不同发展阶段的幼儿并不完全一致，各有侧重点。自闭症的评估需要根据幼儿能力发展的先后顺序，结合环境特点设计测评项目。但个体在幼儿期发展迅速，即使是 3 个月内就能呈现较大差异，若采用同一工具对 6 个月跨度的婴幼儿进行测评，在测评内容、标准单一的情况下，评估的准确性很难保证，即存在测评工具的敏感性和特异性对于不同年龄跨度不稳定的问题(Nah et al., 2019)。为了使婴幼儿自闭症的早期识别更精准，清晰描述自闭症早期患儿的行为特征和发展规律至关重要。研究通常以 3 月龄为间隔对自闭症婴幼儿的早期特征进行追踪和比较(Kaur et al., 2018)，但涉及到的年龄跨度并不完整地涵盖自闭症早期识别的关键时期(即 6~36 个月)，同时追踪时间较长的研究则往往追踪间隔也较大(如 6 个月或 12 个月)且不均匀。同时，不同的研究往往仅关注自闭症的少数典型行为，如物品分享、社会习惯微笑等，并不全面涉及自闭症婴幼儿的所有典型表现。因此，当前国内外仍缺乏对自闭症婴幼儿在研究时间和内容两个维度上更细化且系统的研究，缺乏具有针对性的自闭症早期患儿的识别体系。

第二，缺乏考察自闭症早期患儿情绪、情感能力的测评项目。自闭症早期评估的工具繁多，大多数测评工具可分为社会交往与互动、语言及认知发展、重复性刻板行为/兴趣三方面维度。根据其年龄适用性，各量表关注的侧重点存在差异，一些量表只侧重部分核心症状，如 CHAT 量表关注共同注意和假想游戏能力，仅涉及社会互动方面(尤娜, 杨广学, 2006)。作为诊断“金标准”的 ADOS 能较为全面地观察到自闭症患儿的表现，但其诊断标准中仍然缺少对患儿情绪情感能力发展的评估指标。重要的是，自闭症婴幼儿往往具有情绪和情感发展障碍，在情感淡漠、情绪理解

和表达上存在困难等特征，这些特征通常被认为是社交障碍的典型表现。但是，目前对自闭症婴幼儿社交能力的观察只聚焦在“是否出现互动”的行为层面上，并没有重点突出情绪和情感发展的指标(de Bildt et al., 2015)。情绪情感发展是婴幼儿社会性发展的基础，认识到自闭症早期患儿在情绪情感发展上的特征具有重要意义，可以有针对性地提供干预措施或干预效果评估指标，有助于探索帮助自闭症婴幼儿全面发展的稳健路径。

第三，缺乏现有筛查和诊断方式的创新和有机整合。自闭症的诊断方法主要是医师进行病史分析、症状查询、婴幼儿行为观察和量表测评，同时结合少量 CT/核磁等神经影像检查和遗传学染色体分型检查、代谢病筛查等辅助检查。诊断方式缺乏细致和统一的标准，而且常用方式和工具又存在一定的局限性。例如，依赖于婴幼儿主要养育人的半结构化测评缺少儿童的互动参与，易受到养育人主观因素影响，不能客观反映婴幼儿的能力水平，养育者对于要考察的问题没有精确的理解，可能会高估或低估婴幼儿的能力(Johnson & Myers, 2007)；同时，依赖于现场观察的测评程序十分复杂，且必须让拥有专业资格的医师进行观察才能得到结果(Matson et al., 2011; Romero-Garcia et al., 2019)。国际上以 ADOS 诊断为金标准，但是国内接受过 ADOS 评价的专业人员却极度匮乏，临床医生通过观察得出的结论具有一定的主观性，往往受其临床经验和专业培训影响(Romero-Garcia et al., 2019)，医生与儿童在一起的时间很短暂，难以全面评价婴幼儿在不同情境中的表现，注意到婴幼儿多方面的异常行为(Fitzgerald, 2017; Zabihi et al., 2020)。有研究表明完整的诊断过程会花费相当长的时间(平均 41 个月)和较高的费用，而且量表的评分较为粗略(Hyman et al., 2020)，不仅不利于尽早地对患儿进行诊断，还可能会由于标准不统一而造成误判。

综上，传统筛查与诊断方式受到时间、人力和婴幼儿发育阶段的特殊性等因素的限制，不仅有必要对已有的诊断工具和医生专业诊断标准进行更细致的研究和补充，较为全面地构建自闭症患儿的行为特征体系，以提供对不同年龄段的婴幼儿自闭症的精确识别标准；此外，开发一种能够辅助或者代替医生在家庭或者社区医院进行快速筛查并能够有效减少主观误差的智能工具，将有非常重要的临床价值。

4.2 缺乏融合多模态数据的智能化识别研究

目前，自闭症智能化识别的主要对象群体是 3 岁以上的自闭症儿童和成年人。研究者相较于实验室数据，较少使用临床诊断和自然状态下采集的行为数据。在自闭症智能化识别的 6 个研究子领域内需解决的难题存在一定差异，例如，在基于自闭症面部表情和情绪数据的识别方面，当前已能够实现基于深度学习方法(例如，CNN, DCNN)对自闭症患者的照片或视频帧进行情绪分类(Li et al., 2019; Shukla et al., 2017)。然而，多数情绪分类器的开发仅涉及自闭症成年群体，对自闭症婴幼儿的泛化性较差(Kalantarian et al., 2019)；其次，仅能实现对患者的基本情绪进行定性识别(如，快乐，厌恶，愤怒等)，缺乏对患者情绪复杂性或情绪异常等级的评估研究(Guha et al., 2016)。在基于眼动数据的识别方面，临床和实证研究已累积丰富的自闭症幼儿与典型发展幼儿的眼动模式差异证据，同时领域内已提出有效的特征提取方法、预测框架和评分框架(Liu et al., 2016; Liu et al., 2015)。然而，多数眼动数据是在高度受控的实验室环境中借助专业仪器获得，并要求被试持续注视屏幕，难以在大规模筛查中施行，也很难评估幼儿与抚养者在自然社交互动中的眼神接触。在基于运动控制和运动模式数据的识别方面，目前实现了较为简单的经典任务中固定身体部位和肢体动作的识别，如婴幼儿自闭症诊断中的“叫名反应”的头部运动(Dawson et al., 2018)和抓握动作(Martin et al., 2018)等，还未涉及自闭症幼儿在复杂任务中的肢体动作和身体部位的姿态识别。然而，已有研究大多基于单模态数据进行自闭症患者的智能化识别，尚未充分利用自闭症患者的多模态信息。

当前，自闭症智能化识别的主要趋势是尽可能获取丰富的自闭症婴幼儿的多模态数据，融合多种形态的数据及患者不同的行为维度(如，眼动、语音、唇动、面部表情和肢体动作)，探索不同模态信息间的互补性关系、特征转化与表征模式，采用多模态数据融合思想构建自闭症患儿自动甄别模型，有望实现自闭症婴幼儿筛查方法上的革新和诊断精度的突破。国际上已经开展了大量基于多模态数据的、针对其它精神疾病(比如阿尔茨海默病、精神分裂症)的智能诊断研究，多项

研究表明利用多模态数据能得到较单模态数据更高的诊断精度。这类研究从建模数据来看，主要以 6 岁以上的自闭症患儿为主，对低龄自闭症患儿的自动化识别依然是研究缺口。并且，当前研究者能够获取的自闭症患儿多模态信息较少，主要来自人为控制的实验环境、公开数据集、网络视频资源等，且行为数据大多来自单一任务情境或治疗康复情境，能否有效地应用于大规模的自闭症婴幼儿筛查尚且未知。此外，许多模型识别的自闭症婴幼儿行为标签较为粗略和简单，如基于自闭症观察量表的简单行为标记或基本情绪分类，难以在临床诊断中发挥重要作用。

自闭症智能化识别的基础是构建多模态数据集，没有数据也就不存在发展高精度算法的可能。如何高效率且便捷的获取患儿的多模态信息、减少噪音并进行有效识别是未来研究的重点。目前，国外有较多情境来源的、包含行为标记的、可供研究者使用的自闭症患者公开数据集，尽管还存在着上述的缺乏婴幼儿数据、数据类型较为单一和行为标签粗略等问题，但仍然为自闭症识别算法的快速更新起到了重要的作用。我国罹患自闭症的幼儿数量庞大且逐年递增，但自闭症患儿病例数据分散。要构建智能辅助的自闭症早期筛查、诊断系统，首先应厘清自闭症早期诊断标准，构建我国自闭症婴幼儿异常行为指标评估体系，从多方获取自闭症患儿的多模态数据并进行粗细颗粒度的异常行为标记。组建大规模自闭症及高危婴幼儿数据库和行为特征库是我国开展高质量的智能化识别研究的必要条件。在当前大规模数据集尚未建成的时期，面对模型训练数据需求量大与自闭症患儿样本匮乏之间的矛盾，研究者可先尝试运用深度学习领域的小样本学习(few-shot learning)，如模型微调(fine-tuning)、数据增强(data augmentation)、迁移学习(transfer learning)等方法解决小样本环境下的模型建模和优化问题。

参考文献

毕小彬，范晓壮，米文丽，贺荟中. (2020). 高风险婴儿前瞻性纵向研究与孤独症谱系障碍早期识别. *心理科学进展, 28*(3), 443–455.

陈顺森，白学军，张日昇. (2011). 自闭症谱系障碍的症状、诊断与干预. *心理科学进展, 19*(1), 60–72.

廖梦怡，陈靓影，王广帅，彭世新. (2021). 融合多模态数据的自闭症谱系障碍儿童智能化识别及其有效性. *科学通报, 66*(20), 2618–2628.

徐云，杨健. (2014). 自闭症早期发现研究进展. *中国临床心理学杂志, 22*(6), 1023–1027.

尤娜，杨广学. (2006). 自闭症诊断与干预研究综述. *中国特殊教育, 7*, 26–31.

Ahmed, A. A., & Goodwin, M. S. (2017, May). Automated detection of facial expressions during computer-assisted instruction in individuals on the autism spectrum. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 6050−6055). Association for Computing Machinery.

Akshoomoff, N., Corsello, C., & Schmidt, H. (2006). The role of the autism diagnostic observation schedule in the assessment of autism spectrum disorders in school and community settings. *The California School Psychologist, 11*(1), 7–19.

Anzulewicz, A., Sobota, K., & Delafield-Butt, J. T. (2016). Toward the autism motor signature: Gesture patterns during smart tablet gameplay identify children with autism. *Scientific Reports, 6*(1), 31107.

Arulampalam, M. S., Maskell, S., Gordon, N., & Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing, 50*(2), 174–188.

Baltrušaitis, T., Mahmoud, M., & Robinson, P. (2015, May). Cross-dataset learning and person-specific normalisation for automatic action unit detection. *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)* (Vol. 6, pp. 1–6). IEEE.

Baltrušaitis, T., Robinson, P., & Morency, L. P. (2013). Constrained local neural fields for robust facial landmark detection in the wild. *Proceedings of the IEEE International Conference on Computer Vision Workshops* (pp. 354–361). IEEE.

Baltrušaitis, T., Robinson, P., & Morency, L. P. (2016, March). Openface: An open source facial behavior analysis toolkit. *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 1–10). IEEE.

Baron-Cohen, S., Allen, J., & Gillberg, C. (1992). Can autism be detected at 18 months? The needle, the haystack, and the CHAT. *The British Journal of Psychiatry, 161*(6), 839–843.

Bidwell, J., Essa, I. A., Rozga, A., & Abowd, G. D. (2014, November). Measuring child visual attention using markerless head tracking from color and depth sensing cameras. *Proceedings of the 16th International Conference*

on *Multimodal Interaction* (pp. 447–454). Association for Computing Machinery.

Bovery, M., Dawson, G., Hashemi, J., & Sapiro, G. (2019). A scalable off-the-shelf framework for measuring patterns of attention in young children and its application in autism spectrum disorder. *IEEE Transactions on Affective Computing, 12*(3), 722–731.

Campbell, D. J., Chang, J., Chawarska, K., & Shic, F. (2014, March). Saliency-based Bayesian modeling of dynamic viewing of static scenes. *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 51–58). Association for Computing Machinery.

Campbell, K., Carpenter, K. L., Hashemi, J., Espinosa, S., Marsan, S., Borg, J. S., ... Dawson, G. (2019). Computer vision analysis captures atypical attention in toddlers with autism. *Autism, 23*(3), 619–628.

Chawarska, K., Macari, S., & Shic, F. (2013). Decreased spontaneous attention to social scenes in 6-month-old infants later diagnosed with autism spectrum disorders. *Biological Psychiatry, 74*(3), 195–203.

Chen, S., & Zhao, Q. (2019). Attention-based autism spectrum disorder screening with privileged modality. *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 1181–1190). IEEE.

Chong, E., Chanda, K., Ye, Z., Southerland, A., Ruiz, N., Jones, R. M., ... Rehg, J. M. (2017). Detecting gaze towards eyes in natural social interactions and its use in child assessment. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 1*(3), 1–20. Association for Computing Machinery.

Choutas, V., Weinzaepfel, P., Revaud, J., & Schmid, C. (2018). Potion: Pose motion representation for action recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 7024–7033). IEEE.

Chrysouli, C., Vretos, N., & Daras, P. (2018, September). Affective state recognition based on eye gaze analysis using two–stream convolutional networks. *2018 IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP)* (pp. 1–6). IEEE.

Dawson, G., Campbell, K., Hashemi, J., Lippmann, S. J., Smith, V., Carpenter, K., ... Sapiro, G. (2018). Atypical postural control can be detected via computer vision analysis in toddlers with autism spectrum disorder. *Scientific Reports, 8*(1), 17008.

de Belen, R. A. J., Bednarz, T., Sowmya, A., & Del Favero, D. (2020). Computer vision in autism spectrum disorder research: A systematic review of published studies from 2009 to 2019. *Translational Psychiatry, 10*(1), 1–20.

de Bildt, A., Sytema, S., Zander, E., Bölte, S., Sturm, H., Yirmiya, N., ... Oosterling, I. J. (2015). Autism Diagnostic Interview-Revised (ADI-R) algorithms for toddlers and young preschoolers: Application in a non-US sample of 1, 104 children. *Journal of Autism and Developmental Disorders, 45*(7), 2076–2091.

Dickinson, A., Daniel, M., Marin, A., Gaonkar, B., Dapretto, M., McDonald, N. M., & Jeste, S. (2021). Multivariate neural connectivity patterns in early infancy predict later autism symptoms. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 6*(1), 59–69.

Dietz, C., Swinkels, S., van Daalen, E., van Engeland, H., & Buitelaar, J. K. (2006). Screening for autistic spectrum disorder in children aged 14–15 months. II: Population screening with the Early Screening of Autistic Traits Questionnaire (ESAT). Design and general findings. *Journal of Autism and Developmental Disorders, 36*(6), 713–722.

Elsabbagh, M., Fernandes, J., Webb, S. J., Dawson, G., Charman, T., Johnson, M. H., & British Autism Study of Infant Siblings Team. (2013). Disengagement of visual attention in infancy is associated with emerging autism in toddlerhood. *Biological Psychiatry*, *74*(3), 189–194.

Emerson, R. W., Adams, C., Nishino, T., Hazlett, H. C., Wolff, J. J., Zwaigenbaum, L., ... Piven, J. (2017). Functional neuroimaging of high-risk 6-month-old infants predicts a diagnosis of autism at 24 months of age. *Science Translational Medicine, 9*(393), eaag2882. doi: 10.1126/scitranslmed.aag2882

Eslami, T., Almuqhim, F., Raiker, J. S., & Saeed, F. (2021). Machine learning methods for diagnosing autism spectrum disorder and attention-deficit/hyperactivity disorder using functional and structural MRI: A survey. *Frontiers in Neuroinformatics, 14,* 575999. https://doi.org/10.3389/fninf.2020.575999

Esposito, G., Venuti, P., Maestro, S., & Muratori, F. (2009). An exploration of symmetry in early autism spectrum disorders: Analysis of lying. *Brain and Development, 31*(2), 131–138.

Feil-Seifer, D., & Matarić, M. (2010, March). Using proxemics to evaluate human-robot interaction. *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 143–144). IEEE.

Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM, 24*(6), 381–395.

Fitzgerald, M. (2017). The clinical utility of the ADI-R and ADOS in diagnosing autism. *The British Journal of Psychiatry*, *211*(2), 117–117.

Gabard-Durnam, L. J., Wilkinson, C., Kapur, K., Tager-Flusberg, H., Levin, A. R., & Nelson, C. A. (2019). Longitudinal EEG power in the first postnatal year differentiates autism outcomes. *Nature Communications, 10*(1), 4188.

Gilmore, J. H., Knickmeyer, R. C., & Gao, W. (2018). Imaging structural and functional brain development in early childhood. *Nature Reviews Neuroscience, 19*(3), 123–137.

Girdhar, R., Gkioxari, G., Torresani, L., Paluri, M., & Tran, D. (2018). Detect-and-track: Efficient pose estimation in videos. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 350–359). IEEE.

Guha, T., Yang, Z., Grossman, R. B., & Narayanan, S. S. (2016). A computational study of expressive facial dynamics in children with autism. *IEEE Transactions on Affective Computing, 9*(1), 14–20.

Han, J., Li, X., Xie, L., Liu, J., Wang, F., & Wang, Z. (2018, November). Affective Computing of Childern with Authism Based on Feature Transfer. *2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS)* (pp. 845–849). IEEE.

Hashemi, J., Campbell, K., Carpenter, K., Harris, A., Qiu, Q., Tepper, M., ... Sapiro, G. (2015). A scalable app for measuring autism risk behaviors in young children: A technical validity and feasibility study. *Proceedings of the 5th EAI International Conference on Wireless Mobile Communication and Healthcare* (pp. 23–27). ICST.

Hashemi, J., Dawson, G., Carpenter, K. L., Campbell, K., Qiu, Q., Espinosa, S., ... Sapiro, G. (2018). Computer vision analysis for quantification of autism risk behaviors. *IEEE Transactions on Affective Computing, 12*(1), 215–226.

Hashemi, J., Tepper, M., Vallin Spina, T., Esler, A., Morellas, V., Papanikolopoulos, N., ... Sapiro, G. (2014). Computer vision tools for low-cost and noninvasive measurement of autism-related behaviors in infants. *Autism Research and Treatment, 2014*. https://doi.org/10.1155/2014/935686

Hazlett, H. C., Gu, H., Munsell, B. C., Kim, S. H., Styner, M., Wolff, J. J., ... The IBIS Network (2017). Early brain development in infants at high risk for autism spectrum disorder. *Nature, 542*(7641), 348–351.

Hodges, H., Fealko, C., & Soares, N. (2020). Autism spectrum disorder: definition, epidemiology, causes, and clinical evaluation. *Translational Pediatrics, 9*(Suppl 1), S55–S65. https://doi.org/10.21037/tp.2019.09.09

Hong, S. J., Vogelstein, J. T., Gozzi, A., Bernhardt, B. C., Yeo, B. T., Milham, M. P., & Di Martino, A. (2020). Toward neurosubtypes in autism. *Biological Psychiatry, 88*(1), 111–128.

Hyman, S. L., Levy, S. E., & Myers, S. M. (2020). Identification, evaluation, and management of children with autism spectrum disorder. *Pediatrics, 145*(1), e20193447. https://doi.org/10.1542/peds.2019-3447

Imafuku, M., Hakuno, Y., Uchida-Ota, M., Yamamoto, J.-I., & Minagawa, Y. (2014). "Mom called me!" Behavioral and prefrontal responses of infants to self-names spoken by their mothers. *NeuroImage, 103*, 476–484.

Jeni, L. A., & Cohn, J. F. (2016). Person-independent 3d gaze estimation using face frontalization. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 87–95). IEEE.

Jeni, L. A., Cohn, J. F., & Kanade, T. (2015, May). Dense 3D face alignment from 2D videos in real-time. *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)* (Vol. 1, pp. 1–8). IEEE.

Johnson, C. P., & Myers, S. M. (2007). Identification and evaluation of children with autism spectrum disorders. *Pediatrics, 120*(5), 1183–1215.

Kalantarian, H., Jedoui, K., Washington, P., Tariq, Q., Dunlap, K., Schwartz, J., & Wall, D. P. (2019). Labeling images with facial emotion and the potential for pediatric healthcare. *Artificial Intelligence in Medicine, 98*, 77–86.

Kaur, M., Srinivasan, S. M., & Bhat, A. N. (2018). Comparing motor performance, praxis, coordination, and interpersonal synchrony between children with and without Autism Spectrum Disorder (ASD). *Research in Developmental Disabilities, 72*, 79–95.

Krug, D. A., Arick, J., & Almond, P. (1980). Behavior checklist for identifying severely handicapped individuals with high levels of autistic behavior. *Journal of Child Psychology and Psychiatry, 21*(3), 221–229.

Landa, R., & Garrett‐Mayer, E. (2006). Development in infants with autism spectrum disorders: A prospective study. *Journal of Child Psychology and Psychiatry, 47*(6), 629–638.

Leo, M., Carcagnì, P., Distante, C., Mazzeo, P. L., Spagnolo, P., Levante, A., ... Lecciso, F. (2019). Computational analysis of deep visual data for quantifying facial expression production. *Applied Sciences, 9*(21), 4542. https://doi.org/10.3390/app9214542

Li, B., Mehta, S., Aneja, D., Foster, C., Ventola, P., Shic, F., & Shapiro, L. (2019, September). A facial affect analysis system for autism spectrum disorder. *2019 IEEE International Conference on Image Processing (ICIP)* (pp. 4549–4553). IEEE.

Li, J., & Li, S. (2015). Gaze estimation from color image based

on the eye model with known head pose. *IEEE Transactions on Human-Machine Systems, 46*(3), 414–423.

Liu, W., Li, M., & Yi, L. (2016). Identifying children with autism spectrum disorder based on their face processing abnormality: A machine learning framework. *Autism Research, 9*(8), 888–898.

Liu, W., Yu, X., Raj, B., Yi, L., Zou, X., & Li, M. (2015, September). Efficient autism spectrum disorder prediction with eye movement: A machine learning framework. *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)* (pp. 649–655). IEEE.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 24*(5), 659–685.

Lu, F., Sugano, Y., Okabe, T., & Sato, Y. (2014). Adaptive linear regression for appearance-based gaze estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 36*(10), 2033–2046.

Maenner, M. J., Shaw, K. A., Baio, J., Washington, A., Patrick, M., DiRienzo, M., ... Dietz, P. M. (2020). Prevalence of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2016. *MMWR Surveillance Summaries, 69*(4), 1–12.

Matson, J. L., Wilkins, J., Sevin, J. A., Knight, C., Boisjoli, J. A., & Sharp, B. (2009). Reliability and item content of the Baby and Infant Screen for Children with aUtIsm Traits (BISCUIT): Parts 1–3. *Research in Autism Spectrum Disorders, 3*(2), 336–344.

Martin, K. B., Hammal, Z., Ren, G., Cohn, J. F., Cassell, J., Ogihara, M., ... Messinger, D. S. (2018). Objective measurement of head movement differences in children with and without autism spectrum disorder. *Molecular Autism, 9*(1), 14.

Matson, J. L., & Goldin, R. L. (2014). Diagnosing young children with autism. *International Journal of Developmental Neuroscience, 39*, 44–48.

Matson, J. L., Rieske, R. D., & Tureck, K. (2011). Additional considerations for the early detection and diagnosis of autism: Review of available instruments. *Research in Autism Spectrum Disorders, 5*(4), 1319–1326.

Matson, J. L., Wilkins, J., & Gonzalez, M. (2008). Early identification and diagnosis in autism spectrum disorders in young children and infants: How early is too early? *Research in Autism Spectrum Disorders, 2*(1), 75–84.

Moghadas, M., & Moradi, H. (2018, October). Analyzing human-robot interaction using machine vision for autism screening. *2018 6th RSI International Conference on Robotics and Mechatronics (IcRoM)* (pp. 572–576). IEEE.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Prisma Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Physical Therapy, 89*(9), 873–880.

Nah, Y. H., Young, R. L., & Brewer, N. (2019). Development of a brief version of the Autism Detection in Early Childhood. *Autism, 23*(2), 494–502.

Ozonoff, S., Iosif, A.-M., Baguio, F., Cook, I. C., Hill, M. M., Hutman, T., ... Young, G. S. (2010). A prospective study of the emergence of early behavioral signs of autism. *Journal of the American Academy of Child & Adolescent Psychiatry, 49*(3), 256–266.

Ozonoff, S., Young, G. S., Goldring, S., Greiss-Hess, L., Herrera, A. M., Steele, J., ... Rogers, S. J. (2008). Gross motor development, movement abnormalities, and early identification of autism. *Journal of Autism and Developmental Disorders, 38*(4), 644–656.

Pierce, K., Gazestani, V. H., Bacon, E., Barnes, C. C., Cha, D., Nalabolu, S., ... Courchesne, E. (2019). Evaluation of the diagnostic stability of the early autism spectrum disorder phenotype in the general population starting at 12 months. *JAMA Pediatrics, 173*(6), 578–587.

Poria, S., Cambria, E., Bajpai, R., & Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion, 37*, 98–125.

Rehg, J., Abowd, G., Rozga, A., Romero, M., Clements, M., Sclaroff, S., ... Ye, Z. (2013). Decoding children's social behavior. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3414–3421). IEEE.

Reznick, J. S., Baranek, G. T., Reavis, S., Watson, L. R., & Crais, E. R. (2007). A parent-report instrument for identifying one-year-olds at risk for an eventual diagnosis of autism: the first year inventory. *Journal of Autism and Developmental Disorders, 37*(9), 1691–1710.

Robins, D. L., Fein, D., Barton, M. L., & Green, J. A. (2001). The Modified Checklist for Autism in Toddlers: an initial study investigating the early detection of autism and pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 31*(2), 131–144.

Romero-Garcia, R., Warrier, V., Bullmore, E. T., Baron-Cohen, S., & Bethlehem, R. A. (2019). Synaptic and transcriptionally downregulated genes are associated with cortical thickness differences in autism. *Molecular Psychiatry, 24*(7), 1053–1064.

Scassellati, B. (2007). How social robots will help us to diagnose, treat, and understand autism. *Robotics Research* (pp. 552–563), Springer.

Siegel, B. (2004). *PDDST-II: Pervasive Developmental Disorders Screening Test-II: Early Childhood Screener for Autistic Spectrum Disorder.* Pearson.

Shi, L., Zhou, Y., Ou, J., Gong, J., Wang, S., Cui, X., ... Luo, X. (2015). Different visual preference patterns in response to simple and complex dynamic social stimuli in preschool-aged children with autism spectrum disorders. *PLoS One, 10*(3), e0122280. https://doi.org/10.1371/journal.pone.0122280

Shukla, P., Gupta, T., Saini, A., Singh, P., & Balasubramanian, R. (2017, March). A deep learning frame-work for recognizing developmental disorders. *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 705–714). IEEE.

Stone, W. L., Coonrod, E. E., & Ousley, O. Y. (2000). Brief report: screening tool for autism in two-year-olds (STAT): Development and preliminary data. *Journal of Autism and Developmental Disorders, 30*(6), 607–612.

Syeda, U. H., Zafar, Z., Islam, Z. Z., Tazwar, S. M., Rasna, M. J., Kise, K., & Ahad, M. A. R. (2017, September). Visual face scanning and emotion perception analysis between autistic and typically developing children. *Proceedings of the 2017 Acm International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 Acm International Symposium on Wearable Computers* (pp. 844–853). Association for Computing Machinery.

Talbott, M. R., & Miller, M. R. (2020). Future directions for infant identification and intervention for autism spectrum disorder from a transdiagnostic perspective. *Journal of Clinical Child & Adolescent Psychology, 49*(5), 688–700.

Tang, C., Zheng, W., Zong, Y., Cui, Z., Qiu, N., Yan, S., & Ke, X. (2018, October). Automatic smile detection of infants in mother-infant interaction via CNN-based feature learning. *Proceedings of the Joint Workshop of the 4th Workshop on Affective Social Multimedia Computing and first Multi-Modal Affective Computing of Large-Scale Multimedia Data* (pp. 35–40). Association for Computing Machinery.

Taylor, L. J., Eapen, V., Maybery, M., Midford, S., Paynter, J., Quarmby, L., ... Whitehouse, A. J. (2017). Brief report: An exploratory study of the diagnostic reliability for autism spectrum disorder. *Journal of Autism and Developmental Disorders, 47*(5), 1551–1558.

Teitelbaum, P., Teitelbaum, O., Nye, J., Fryman, J., & Maurer, R. G. (1998). Movement analysis in infancy may be useful for early diagnosis of autism. *Proceedings of the National Academy of Sciences, 95*(23), 13982–13987.

Thevenot, J., López, M. B., & Hadid, A. (2017). A survey on computer vision for assistive medical diagnosis from faces. *IEEE Journal of Biomedical and Health Informatics, 22*(5), 1497–1511.

Vyas, K., Ma, R., Rezaei, B., Liu, S., Neubauer, M., Ploetz, T., ... Ostadabbas, S. (2019, October). Recognition of atypical behavior in autism diagnosis from video using pose estimation over time. *2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)* (pp. 1–6). IEEE.

Wang, Z., Cai, H., & Liu, H. (2018, December). Robust eye center localization based on an improved SVR method. *International Conference on Neural Information Processing* (pp. 623–634). Springer, Cham.

Wang, Z., Liu, J., He, K., Xu, Q., Xu, X., & Liu, H. (2019). Screening early children with autism spectrum disorder via response-to-name protocol. *IEEE Transactions on Industrial Informatics, 17*(1), 587–595.

Wiggins, L. D., Baio, J., & Rice, C. (2006). Examination of the time between first evaluation and first autism spectrum diagnosis in a population-based sample. *Journal of Developmental & Behavioral Pediatrics, 27*(2), S79–S87.

Winoto, P., Chen, C. G., & Tang, T. Y. (2016, September). The development of a Kinect-based online socio-meter for users with social and communication skill impairments: A computational sensing approach. *2016 IEEE International Conference on Knowledge Engineering and Applications (ICKEA)* (pp. 139–143). IEEE.

Wolfers, T., Floris, D. L., Dinga, R., van Rooij, D., Isakoglou, C., Kia, S. M., ... Beckmann, C. F. (2019). From pattern classification to stratification: Towards conceptualizing the heterogeneity of Autism Spectrum Disorder. *Neuroscience & Biobehavioral Reviews, 104*, 240–254.

Xie, X., Niu, J., Liu, X., Chen, Z., Tang, S., & Yu, S. (2021). A survey on incorporating domain knowledge into deep learning for medical image analysis. *Medical Image Analysis, 69*. https://doi.org/10.1016/j.media.2021.101985

Xu, L., Liu, Y., Yu, J., Li, X., Yu, X., Cheng, H., & Li, J. (2020). Characterizing autism spectrum disorder by deep learning spontaneous brain activity from functional near-infrared spectroscopy. *Journal of Neuroscience Methods, 331*, 108538.

Young, R. L. (2007). *Autism detection in early childhood: ADEC.* Victoria, Australia: Australian Council of Educational Research.

Zabihi, M., Floris, D. L., Kia, S. M., Wolfers, T., Tillmann, J., Arenas, A. L., ... Marquand, A. (2020). Fractionating autism based on neuroanatomical normative modeling. *Translational Psychiatry, 10*(1), 384.

Zheng, Z., Young, E. M., Swanson, A. R., Weitlauf, A. S.,

Warren, Z. E., & Sarkar, N. (2015). Robot-mediated imitation skill training for children with autism. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 24*(6), 682–691.

Zhou, H., Xu, X., Yan, W., Zou, X., Wu, L., Luo, X., ... Wang, Y. (2020). Prevalence of autism spectrum disorder in China: a nationwide multi-center population-based study among children aged 6 to 12 years. *Neuroscience Bulletin, 36*(9), 961–971.

Zunino, A., Morerio, P., Cavallo, A., Ansuini, C., Podda, J., Battaglia, F., ... Murino, V. (2018, August). Video gesture analysis for autism spectrum disorder detection. *2018 24th International Conference on Pattern Recognition (ICPR)* (pp. 3421–3426). IEEE.

Zwaigenbaum, L., Bryson, S., Rogers, T., Roberts, W., Brian, J., & Szatmari, P. (2005). Behavioral manifestations of autism in the first year of life. *International Journal of Developmental Neuroscience, 23*(2-3), 143–152.

Early screening and diagnosis of autism spectrum disorder assisted by artificial intelligence

YUAN Yuzhuo[1], LUO Fang[2]

([1] *Collaborative Innovation Center of Assessment toward Basic Education Quality, Beijing Normal University, Beijing* 100875, *China*) ([2] *Faculty of Psychology, Beijing Normal University, Beijing* 100875, *China*)

Abstract: Symptoms of Autistic Spectrum Disorders (ASD) manifest as early as infancy, and the earlier detection and intervention can lead to better therapeutic results. The traditional tools of early screening and diagnosis of autism have limitations in evaluation methods and procedures, which cannot meet the needs of large-scale screening and diagnosis. With the rapid artificial intelligence technological advancement, using an intelligent approach for large-scale non-inductive early screening and diagnosis of autism has become possible. In the past decade, a myriad of research findings on intelligent detection technology of autism were generated domestically and internationally in six aspects: behaviors in classic tasks, facial expressions and emotions, eye gaze data, brain imaging, motor control and movement patterns, and multimodal data. Future research should focus on constructing a domestic intelligent medical screening and diagnosis system for early autism, developing screening tools for infants and young children, constructing an automated recognition model for autistic infants by integrating multimodal data, establishing a refined autism diagnosis method combined with brain imaging technology, and other aspects.

Key words: autism spectrum disorder, early screening and diagnosis of autism, intelligent screening of autism, artificial intelligence, multimodal data